



The data management plan and its four FAIR principles

*Workshop Data Management Plans: FAIRify your Data!
in Weimar on July 1, 2020*

Kerstin Helbig

Humboldt-Universität zu Berlin, Computer and Media Service

<https://orcid.org/0000-0002-2775-6751>





Reasons for good data management

Personal interests

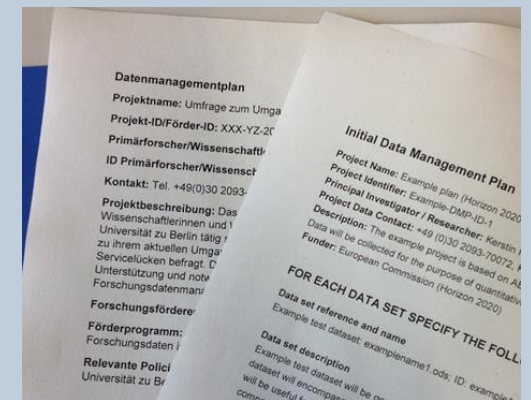
- Good scientific practice
- Knowledge management/knowledge transfer
- Protection against data loss

External interests

- Research funders
- Publishers
- Guidelines of professional associations and university management

What are data management plans?

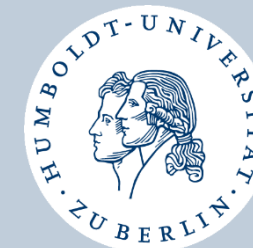
- All information that adequately describes and documents the collection, processing, storage, archiving and publication of research data within a research project
- Analysis of the workflow from data creation to its use*
- Size can vary between a few paragraphs and several pages



* J. Ludwig, H. Enke (Hrsg.): Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt. Verlag Werner Hülsbusch: Glückstadt, 2013.

Photo: Creative Commons CC0

Funder policies



Funder	Plan requirement	Submission with application	Content	Update
European Commission (Horizon 2020)	Data management plan	No, initial plan within first 6 project month	Content of Horizon 2020 template	Update if significant changes arise as well as to project end
German Research Foundation (DFG)	Statement on research data handling	Yes	Content of the Guidelines on the Handling of Research Data	No
Federal Ministry of Education and Research (BMBF)	Plan sometimes required, depending on specific program	If required, yes	Content depends on specific program; Educational research: Content of checklist	Depending on specific program
VolkswagenStiftung	Data management plan	Yes	Content of Science Europe template	No



FAIR Principles

- **F**indable
Research data should be easy to find.
- **A**ccessible
Research data should be easy to access.
- **I**nteroperable
Research data should be interoperable with other systems.
- **R**e-usable
Research data should be reusable.

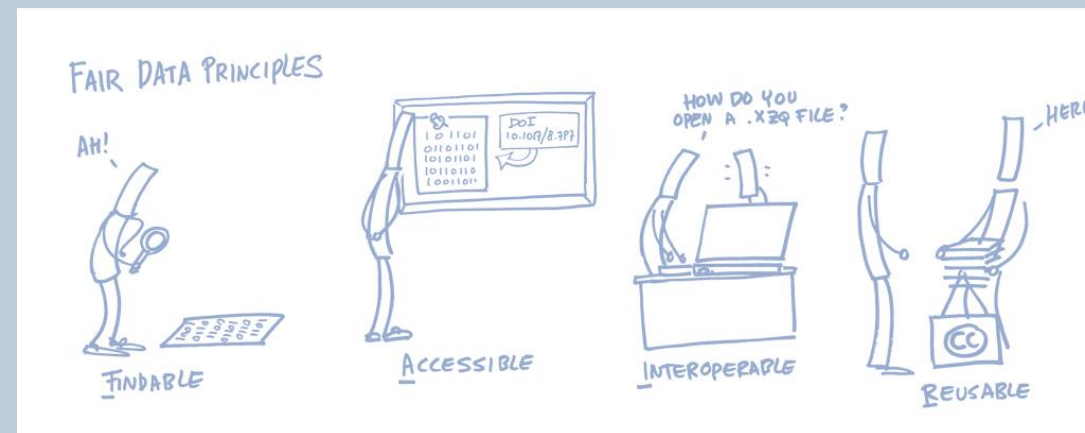


Image: Creative Commons CC0

FAIR Principles:
<https://www.go-fair.org/fair-principles/>



Content of a DMP

- Administrative information (project name, author, contributors, contact, funding program, etc.)
- Project and dataset description
- Information on metadata and standards
- Data sharing
- Archiving and backup
- Responsibilities
- Costs

N.B.: Keep it short and simple, stay realistic.
As informative as possible, as detailed as necessary!



Findable: Data repository

The screenshot shows the re3data.org website interface. On the left is a 'Filter' sidebar with various categories like Subjects, Content Types, Countries, etc. The main area displays search results for 2332 items. Three results are visible:

- CancerData.org**: Sharing data for cancer research. Subject(s): Basic Biological and Medical Research, Medicine, Biology, Life Sciences. Content type(s): Standard office documents, Databases, Images, Structured graphics, Scientific and statistical data formats, Raw data, Plain text, Archived data, other. Country: Netherlands. Description: The CancerData site is an effort of the Medical Informatics and Knowledge Engineering team (MIKE for short) of Maastric Clinic, Maastricht, The Netherlands. Our activities in the field of medical image analysis and data modelling are visible in a number of projects we are running. CancerData is offering several datasets. They are grouped in collections and can be public or private. You can search for public datasets in the NBIA (National Biomedical Imaging Archive) image archives without logging in.
- Claremont Colleges Digital Library (CCDL)**: Subject(s): Humanities, Social and Behavioural Sciences, Social Sciences, Geosciences (including Geography), Geography, Chemistry, Biology, Mathematics, Computer Science, Humanities and Social Sciences, Natural Sciences, Life Sciences, Computer Science, Electrical and System Engineering, Engineering Sciences. Content type(s): Audiovisual data, Images, Plain text, other. Country: United States. Description: The Claremont Colleges Digital Library (CCDL) provides access to historical and visual resources collections created both by and for The Claremont Colleges community.
- PRISM: University of Calgary Institutional Repository**: DSpace at University of Calgary. Subject(s): Humanities, Life Sciences, Natural Sciences, Engineering Sciences, Humanities and Social Sciences. Content type(s): Standard office documents, Images, Scientific and statistical data formats. Country: Canada. Description: PRISM is a digital archive of the university's intellectual output. Established and maintained by Libraries and Cultural Resources to manage, preserve and make available the academic works.

<https://www.re3data.org>





Findable: Choosing a repository



Does the repository have a certificate (e.g. CoreTrustSeal)?



Are persistent identifiers assigned (e.g. DOI, handle)?



How is access to the data (open, restricted, inaccessible)?



Are the terms of use and license of the data mentioned by the repository?

Icons: <https://www.re3data.org>



If you did not find an appropriate data repository...

Generic repositories:

- Zenodo, <https://zenodo.org>
- Figshare, <https://figshare.com>
- DRYAD (life sciences), <http://datadryad.org>

Institutional/regional repository:

- Digital Library Thuringia (DBT), <https://www.db-thueringen.de>



Accessible: Metadata creation

Metadata serve primarily to **find** the data (e.g. creator, time, location).

A discipline-specific **overview** of metadata standards can be found at <http://www.dcc.ac.uk/resources/metadata-standards> and <http://rd-alliance.github.io/metadata-directory/subjects>

- Examples:
 - Dublin Core (generic)
 - ISO 19115 (geosciences)
 - Darwin Core (biodiversity)
 - Data Documentation Initiative (social sciences)
 - TEI Guidelines (linguistics and literature)



Photo: Creative Commons CC0



Example metadata

Determinants of Unemployment in the European Union. An empirical Study of the Federal Republic of Germany (FRG), France, Great Britain and Italy		gesis
DOI	10.4232/1.8198	
Version	1.0.0	
Resource Type	Dataset	
Creator	• Hubert, Frank	
Publication Date	2006	
Language	German	
Classification	• ZA: ◦ Historical Studies Data	
Description	• Abstract Since the oil price shock in 1974 unemployment increased significantly and also did not really decline in periods of economic ups ... see more	

Temporal Coverage	• 1961 - 1993
Geographic Coverage	• Germany (DE) • France (FR) • United Kingdom (GB) • Italy (IT)
Collection Mode	Types of sources: Scientific publications, Statistics published by the OECD Statistics published by the European commission Annual expert ... see more
Data and File Information	• Type of Units: Text Unit Number of Units: 37 Number of Variables: 111
Availability	● Download A - Data and documents are released for academic research and teaching.
Rights	All metadata from GESIS DBK are available free of restriction under the Creative Commons CC0 1.0 Universal Public Domain Dedication. Howeve ... see more

Source: Hubert, Frank (2006): Determinants of Unemployment in the European Union. An empirical Study of the Federal Republic of Germany (FRG), France, Great Britain and Italy. GESIS Data Archive, Cologne. ZA8198 Data file Version 1.0.0, <http://doi.org/10.4232/1.8198>
Image: <https://www.da-ra.de>



Interoperable: File formats

For long-time archiving files should be :

- unencrypted, uncompressed, non-proprietary/patent-encumbered and in an open, documented standard

File type	Recommendation	Avoid
Tabular data	CSV, TSV, SPSS portable	XLS
Text	TXT, ODT, HTML, RTF; PDF/A only if layout is important	DOC, PPT
Multimedia	Container: MP4, Ogg Codec: Theora, Dirac, FLAC	QuickTime, H264
Image	TIFF, JPEG2000, PNG	GIF, JPG
Structured data	XML, RDF, JSON	RDBMS

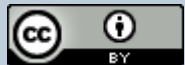
Save files in a recommended format in addition to the original software format!



Re-usable: License



CC0 (Public Domain)



CC BY (Attribution)

CC licenses are not suitable for software. For this purpose, use e.g.:

- MIT license
- GNU General Public Licence (GPL)
- GNU Lesser General Public License (LGPL)
- Apache license

Further information: <http://forschungslizenzen.de> and <https://opensource.org>

Source: <https://creativecommons.org/licenses/>





Re-usable: Documentation

Ways to document your data:

- Laboratory/field diary
- Codebook
- Readme file
- Article in a data journal
- ...



Example: Readme file

- Documents the research process and the data
- Includes:
 - Abstract
 - Hypotheses
 - Information on the collection of data (methods, units, time, place, devices)
 - Measures for data cleansing (deletion of outliers, weighting)
 - Structure of the data and its relationships to each other
 - Explanation of variables, labels and codes
 - Differences between different versions
 - Information on access and terms of use



Practical help

Tools for creating a data management plan:

- DMPonline
<https://dmponline.dcc.ac.uk>
- DMPTool
<https://dmptool.org>
- RDMO (Research Data Management Organiser)
<http://rdmorganiser.github.io>

Further information, templates and sample plans:

<https://www.cms.hu-berlin.de/en/dl-en/dataman-en/work/create-dmp>



DMP examples

There are more and more publicly available data management plans:

- [DMPs in RIOjournal](#)
- [DMPs on Zenodo](#)
- [Example DMPs at the website of the Digital Curation Centre \(DCC\)](#)
- [Publicly available DMPs of DMPTool](#)
- [Annotated collection of DMPs by LIBER](#)



Many thanks for your interest!

Kerstin Helbig, research data management coordinator
kerstin.helbig@cms.hu-berlin.de

Further information on research data management:
<https://hu.berlin/dataman>



Follow us on Twitter [@DatawomenHUB](https://twitter.com/DatawomenHUB)

